(12)

Technical Report 615

Command and Control Training in the Combined Arms Tactical Training Simulator

Gary S. Thomas, Ira T. Kaplan,
and Herbert F. Barber

ARI Field Unit at Fort Leavenworth, Kansas
Systems Research Laboratory

U. S. Army

Research Institute for the Behavioral and Social Sciences

February 1984

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

L. NEALE COSBY
Colonel, IN
Commander

Technical review by

Robert E. Solick
Stanley M. Halpin

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-TST, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report 615	2. GOVT ACCESSION NO. AD-A142 742	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMMAND AND CONTROL TRAINING IN THE COMBINED ARMS TACTICAL TRAINING SIMULATOR		5. TYPE OF REPORT & PERIOD COVERED --
		6. PERFORMING ORG. REPORT NUMBER --
7. AUTHOR(s) Gary S. Thomas Ira T. Kaplan Herbert F. Barber		8. CONTRACT OR GRANT NUMBER(s) --
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue, Alexandria, VA 22333		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q263744A795
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue, Alexandria, VA 22333		12. REPORT DATE February 1984 13. NUMBER OF PAGES 38
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) --		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

--

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Battle simulation	Command group training
Command and control training	Training effectiveness
Command group performance measurement	Training feedback

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Army Training Battle Simulation System (ARTBASS) is now being developed to succeed the Combined Arms Tactical Training Simulator (CATTS) as the Army's most advanced command group training system. The present research assessed the effects on command group performance of the CATTS/ARTBASS technology, supplemented by an Army Research Institute-developed diagnostic and feedback package. Measures of performance included information flow during planning and execution, simulation battlefield outcomes, and ratings of critical tasks. All performance measures increased significantly. (Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
i SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Item 20 (Continued)

from the first (pretest) to the fourth (post-test) exercise day. In a previous experiment using CATTS without added feedback, there was less improvement in performance. It was concluded that CATTS/ARTBASS training is effective when it includes detailed diagnostic feedback.

Technical Report 615

Command and Control Training in the Combined Arms Tactical Training Simulator

Gary S. Thomas, Ira T. Kaplan,
and Herbert F. Barber

Submitted by
Stanley M. Halpin, Chief
ARI Field Unit at Fort Leavenworth, Kansas

Approved as technically adequate
and submitted for publication by
Jerrold M. Levine, Director
Systems Research Laboratory

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

February 1984

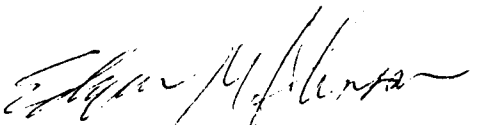
Army Project Number
2Q263744A795

Training and Simulation

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The U.S. Army Research Institute, Fort Leavenworth Field Unit conducts a systems and training research program in support of the Combined Arms Center. This report describes the second of two experiments conducted by the Field Unit contributing to a Training Development Study (TDS) of the Army Training Battle Simulation System (ARTBASS). The first experiment investigated the effects of certain system and scenario characteristics on measures of command group performance. The present experiment measured the effects of command group training supported by ARI-designed diagnostic and feedback procedures. Both experiments were performed in cooperation with the Battle Simulations Directorate at Fort Leavenworth.



EDGAR M. JOHNSON
Technical Director

COMMAND AND CONTROL TRAINING
IN THE COMBINED ARMS TACTICAL TRAINING SIMULATOR

Executive Summary

Objective:

The Combined Arms Center at Fort Leavenworth has responsibility for the development of command and control (C^2) training systems. One of the most promising recent developments is the application of automation to C^2 training systems. Currently, the most advanced of these systems is the Combined Arms Tactical Training Simulator (CATTS), which has been used as a training vehicle and training development test bed. Based on information gathered during the use and refinement of CATTS, a follow-on system, the Army Training Battle Simulation System (ARTBASS), is now being developed. In order to demonstrate that ARTBASS will be effective for training battalion command groups in C^2 behaviors, the ARTBASS Test Integration Working Group decided that a Training Development Study (TDS) could be conducted using CATTS in lieu of ARTBASS due to the similarity between the systems.

In support of CATTS/ARTBASS development and the TDS, the efforts of the ARI Field Unit at Fort Leavenworth have included the refinement of C^2 measurement techniques and the identification of system and scenario characteristics (e.g., weather, combat ratio) which impact on command group C^2 performance in CATTS. The present experiment assessed the effects of CATTS training, supplemented by an ARI-developed diagnostic and feedback package, on the C^2 performance of battalion command groups (BCGs).

Procedures:

Five battalion command groups (players), from three mechanized infantry and two armor battalions, each participated in four one-day CATTS exercises between December 1982 and March 1983. The first and last exercises were designated as pretest and post-test exercises, respectively, and included covering force/delay missions on different portions of Fulda Gap terrain. The intervening training exercises were delay and movement to contact conducted on Ft. Irwin or Sinai terrain. Measures of performance included (1) information reception and transmission by group members during planning (information flow questionnaire), (2) information exchange by staff members during battle execution (probes), (3) degree of success on the simulated

battlefield (mission accomplishment scores), and (4) BCG ARTEP (Army Training and Evaluation Program) performance as assessed by CATTS controllers, player-controllers, and the players themselves (subjective ratings).

Measures of performance were collected on the pretest exercises and, with the exception of mission accomplishment scores, these results were presented to command group members in individual feedback sessions conducted by CATTS controllers and ARI personnel prior to the training exercises. Those areas diagnosed as needing improvement were addressed by the players during the training exercises. At the conclusion of pretest and training exercises, significant events from each day's battle were presented to the players via an "instant replay" capability of the CATTS computer. A CATTS controller conducted the battle replay sessions so that strengths and weaknesses of battle procedures could be identified. Players and player-controllers also conducted after-action reviews. Performance measures were again collected on the post-test exercises.

Findings:

All measures of performance increased significantly from pre- to post-test exercises, thereby demonstrating the training potential of the CATTS/ARTBASS technology. Battalion command group performance on information flow during both planning and execution, subjective ratings of ARTEP performance, and performance on the battlefield all improved with CATTS training. In previous research (Thomas, Barber, and Kaplan; 1983) when detailed feedback was not provided, BCG performance on information flow during planning and on the simulated battlefield did not increase as a result of CATTS training. In addition, the increase in subjective ratings of ARTEP performance observed in the current research was generally greater than in the previous research.

Utilization of Findings:

The degree to which these findings are generalizable to ARTBASS is limited by the degree to which ARTBASS resembles CATTS, as it was used in the current research. Of great importance is the implication that without a mechanism for detailed command group performance diagnosis and feedback, ARTBASS may be less than optimally effective for training BCGs in the field.

COMMAND AND CONTROL TRAINING
IN THE COMBINED ARMS TACTICAL TRAINING SIMULATOR

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIXES

COMMAND AND CONTROL TRAINING IN THE COMBINED ARMS TACTICAL TRAINING SIMULATOR

INTRODUCTION

The advent of automation has generated new and expanded possibilities for the development of training vehicles to meet the Army's needs in the 1980's and 90's. One of the most promising uses of automation is in the area of command and control (C^2) training. Specifically, the Combined Arms Tactical Training Simulator (CATTS), a computer driven, free play C^2, battle simulation system was developed as a training vehicle and training development test bed. Based on information gathered during the development and refinement of CATTS, a follow-on system, the Army Training Battle Simulation System (ARTBASS), is now being developed.

Prior to fielding, it is necessary to demonstrate that ARTBASS is, in fact, effective for training command groups in C^2 behaviors. The development schedule for ARTBASS severely limits the amount of time available to determine its training effectiveness, but since ARTBASS in essence grew from CATTS, there is a close similarity between the two. Therefore, the ARTBASS Test Integration Working Group decided that a Training Development Study (TDS) should be conducted using CATTS in lieu of ARTBASS. The generalizability of TDS results to ARTBASS is, of course, limited by the degree to which the two systems are actually similar.

Previous research (Thomas, Barber, and Kaplan, 1983) identified several system and scenario characteristics (e.g., combat ratio, mission type, weather, etc.) which significantly influenced the success achieved by battalion command groups (BCGs) on the simulated battlefield, as measured by battle outcomes. In that research, performance feedback to BCGs was limited to an after-action review conducted by the players themselves at the conclusion of each exercise day. Neither measures of battlefield performance, nor measures of inter- and intra-staff communication during planning increased with CATTS training. Subjective ratings of BCG performance on a set of ARTEP subtasks provided by company commmanders did not increase as a function of CATTS training. On the other hand, ratings by CATTS controllers and the players themselves increased from the first day to the third and fourth day's exercises.

In support of ARTBASS development and the TDS, the efforts of the ARI Field Unit at Fort Leavenworth included the refinement of C^2 measurement techniques and the development of a diagnostic and feedback module, so that changes in C^2 performance as a function of exposure to CATTS could be documented. The present research assessed the effect of CATTS training, supplemented by performance diagnosis and feedback, on a variety of BCG C^2 measures.

METHOD

Participants

Five battalion command groups each participated in four one-day CATTS training exercises between December 1982 and March 1983. Two of the command groups (players) were from armor battalions, and three were from mechanized infantry battalions. Table 1 characterizes the typical composition of each command group, where each key position (e.g., battalion commander, S1, S2, S3, and S4) was occupied by the individual who normally filled that position, i.e., incumbant. Several supporting members (not controlled) also participated in the exercises.

Experimental Design

A pretest, training, post-test design was used to assess the combined effects of performance feedback and CATTS training on several measures of performance. Performance measures were collected on pre- and post-test exercises, where BCGs conducted covering force/delay missions on two similar and adjacent parts of Fulda Gap terrain. The portions of terrain used were partially counter-balanced, so that three BCG's operated on the northern portion on the pretest and the southern portion on the post-test, and the order of terrain presentation was reversed for the remaining two BCGs. Other potentially important variables such as initial combat ratio, weather and extent of jamming were equivalent for each exercise. Performance feedback on the measures collected during the pretest was presented to the BCGs on the morning after the pretest exercise. Prior to the post-test exercise, each BCG conducted two training exercises where delay and movement to contact missions were executed on desert terrain. At the conclusion of each exercise day there was an after-action review and a replay of the battle conducted by a CATTS controller. (The schedule of events for the exercises appears in Table 2).

Training System

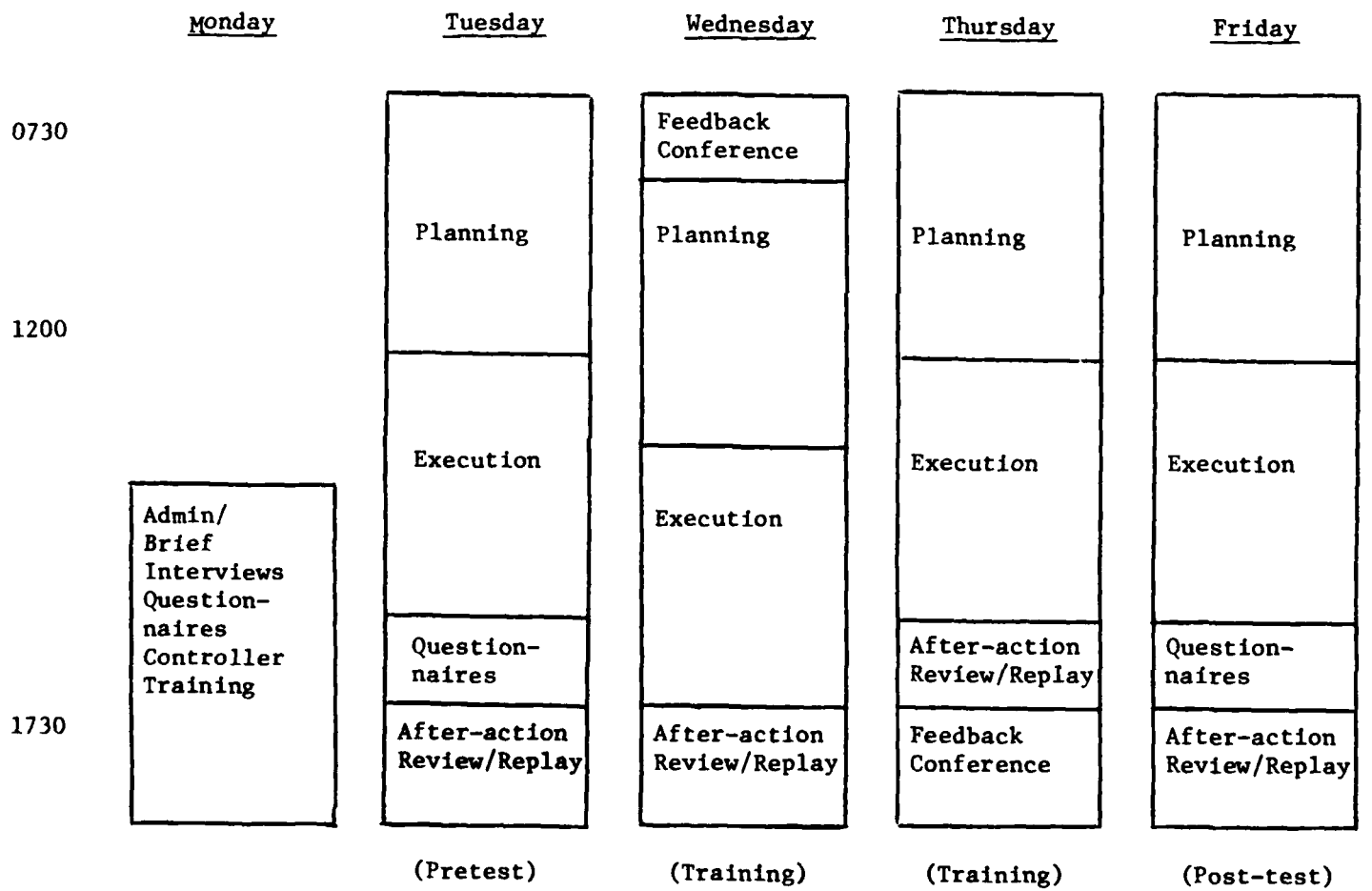
Simulation. The battlefield environment was simulated by the CATTS system, which provided a computer-driven exercise to train maneuver battalion commanders and their staffs in the control and coordination of combined arms operations. CATTS simulated the actions of units in combat, moved elements on and about the battlefield, calculated intervisibility and detection between forces, calculated weapon-to-target ranges, and the effects of weapons employment. It also maintained the status of personnel, equipment, ammunition, and fuel for friendly and enemy forces. Speed of maneuver, line of sight, and weapons effects were affected by changes in weather, terrain contour, soil type, suppressive fires, and personnel and equipment status. Given line of sight, engagements among maneuver weapon systems were automatic.

Table 1

Battalion Command Group Members
Battalion Commander
S1
S4
S1 or S4 NCO
S1 or S4 RTO
S2
S2 NCO
S3
S3 Air
S3 NCO
S3 RTO
Company Commanders (2 tank, 2 line)
Fire Support Officer
Fire Support NCO
Fist Chief (two)
Air Liaison Office (Air Force)
Forward Air Controller (Air Force)

Table 2

General Schedule of Events



The CATTS exercises were conducted in a real-time, free-play mode. Within the prescribed tactical situation, the battalion commander could employ his assets in any manner he deemed appropriate. The only constraints were the assets available to the battalion and the actions of the enemy commander. Deployment of enemy assets was consistent across exercises, but, in accordance with threat doctrine, the threat controller made minor tactical adjustments to counter unique situations created by friendly force operations.

In this research, the command group, except the S1 and S4, occupied a simulated tactical operations center (TOC); the S1 and S4 were in another area, designated as the combat trains. The players (the battalion command group) in both areas were provided with communications equipment normally found in a maneuver battalion. They could communicate with higher, lower, and adjacent units (played by controllers) in any manner consistent with Army procedure and with the simulated location of the various units: face-to-face, by telephone, by radio, and by written message.

Figure 1 illustrates the communication among the players, the controllers, and the computer. Most communication took place by radio and telephone. The BCGs had seven radio nets (actually hard-wired) with appropriate alternate frequencies. The nets included the following: the brigade command, brigade intelligence, brigade administration/logistics, battalion command, battalion administration/logistics, and air support nets. In addition, the command group also had a RATT (radioteletype) unit and field telephones, when appropriate. The sounds of enemy jamming, battle, and engine and generator noise were generated during the exercise to enhance the realism of the experience.

Controllers. A team of controllers, permanently assigned to CATTS, mediated between the players and the computer. The control team consisted of a chief controller, who also played the role of brigade commander, and brigade S1, S2, and S3, and S4 controllers. In addition, a fire support controller, a direct air support controller, and a threat controller were also present. Three additional controllers identified as interactors, input orders into the computer at three control consoles: (1) the command and control interactor input orders from the battalion command group via company commanders to the maneuver units modeled in the computer, (2) the fire support interactor input orders to the friendly artillery and air support units, and (3) the threat interactor input actions directed by the threat controller. All controllers were in the computer control room.

Player-Controllers. Each command group brought along its company commanders and fire support representatives to serve as player-controllers. They received orders from battalion and translated them into subordinate unit maneuvers for input into the computer by the interactors. In addition, they also received battle status reports from the computer and relayed that information back to the command group in the form of situation reports and spot reports. All player-controllers were in the computer control room.

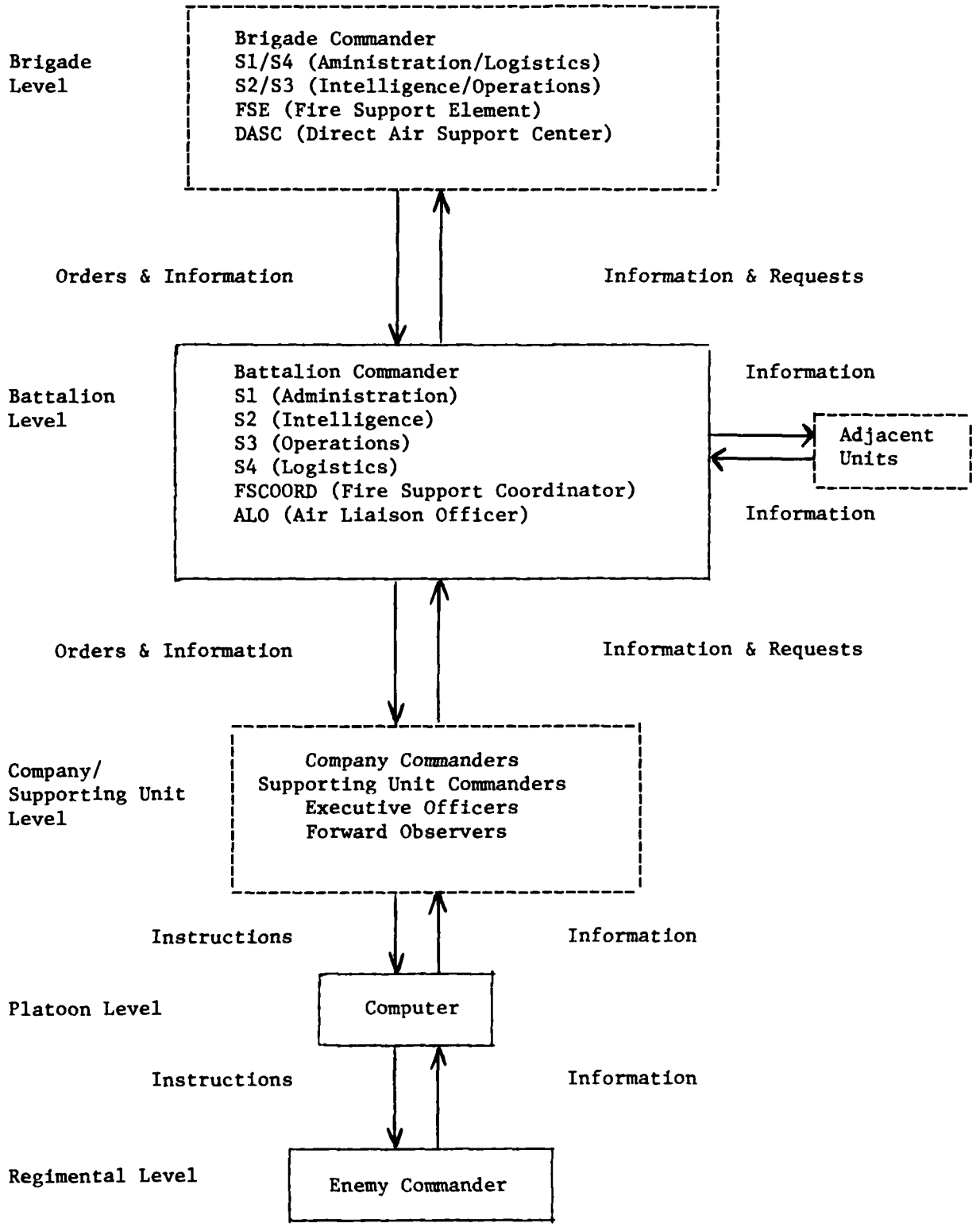


Figure 1. Communication between controller and player positions in CATTS. Controller positions are inclosed by broken lines.

Feedback System

In the battle replay sessions at the end of each exercise day, CATTS controllers replayed significant events on a video display of the terrain, which showed the position and strength of friendly and enemy forces during the battle. Then the players and player-controllers presented their perceptions of the battle in after-action reviews. Feedback conferences were held on the morning following the pretest exercise. ARI personnel provided the battalion commander with results from an information flow questionnaire, described below, and with ratings of BCG performance on a short list of ARTEP tasks. Problem areas were identified and solutions were recommended. The remaining staff members (S1, S2, S3, S4, FSO, and ALO) met with their brigade controller counterparts and discussed problems in command and control processes identified in the pretest exercise. Discussions included, but were not confined to, performance on ARTEP tasks and the group's ability to transmit important information during mission execution (probes). Problem areas were identified and tentative solutions recommended by controllers. At the conclusion of the third exercise day, feedback conferences were again conducted with the same individuals. Discussions focused on problems that were resolved, new problem areas, and recommended solutions to problems.

Performance Measures

The following measures were intended to assess the command group's ability to transmit important information during planning (information flow questionnaire) and execution (probes), to perform ARTEP tasks (subjective ratings), and to perform on the simulated battlefield (mission accomplishment scores).

Information Flow. At the beginning of each pretest and post-test exercise, the principal members of the battalion staff, the S1, S2, S3, S4, FSO, and ALO were briefed separately by their brigade counterparts. In these briefings, certain unique items of information were presented to each member. Then the battalion commander and his staff worked together for three to four hours to develop a plan which they presented to their company commanders during the battalion operations order briefing. Subsequently, the command group members and company commanders answered multiple-choice questions based upon the unique information originally presented in the brigade briefing. Examples of questions are presented in Appendix A. There were two parallel forms of the questionnaire, one for the pretest and one for the post-test. The order of presentation of the questionnaires was counterbalanced across the five units. The responses were analyzed to provide three measures of information flow:

1. Reception of required information that was presented to the staff member during the brigade briefing. The percent of these items answered correctly measured communication from brigade to battalion.

2. Reception of required information that the command group members should have received indirectly from those members who received it directly from brigade. The percent of such items answered correctly measured communication within the battalion command group.

3. Reception of information required by the company commanders that should have been transmitted to them by members of the battalion command group. The percent of such items answered correctly measured communication to the company commanders during the battalion operations order briefing.

Probes. In a procedure similar to the information flow technique, a series of preplanned "probes" were inserted into the exercise during the execution phase. The probes were designed to assess coordination, communication, and information-processing behaviors within the command group. Probes were events that could realistically occur during an operation. Some probes were used in both the pretest and the post-test, because they occurred as a natural outgrowth of the exercise, e.g., asking for a situation report or an estimate of enemy intentions; others were used only once in either the pretest or the post-test, because they might seem artificial if they were repeated, e.g., indications that an adjacent unit was suffering a chemical attack. The unique probes were counterbalanced across command groups to control for difficulty. The number of probes was limited, so as not to interfere with the controllers' primary functions or the flow of the exercise. Each of five brigade controllers (S1, S2, S3, S4, and FSO) inserted four to six probes in each pre- or post-test exercise and evaluated the timeliness, completeness, and accuracy of the response. An example of a probe appears in Appendix B.

Mission Accomplishment Scores. Procedures for deriving mission accomplishment scores that are based on BCG achievement of the objectives of a covering force/delay mission are described in Thomas and Cocklin (1983). That research described how performance on separate mission objectives could be combined into an overall assessment of battlefield mission accomplishment. The current research used the same objectives of a covering force/delay mission as did the previous work. These measures included (1) relative losses between opposing forces (surviving maneuver force ratio differential), (2) whether or not friendly forces were combat effective (50% of initial strength) at the conclusion of battle, (3) the quality of friendly intelligence gathering (estimates of enemy strength, location, rate of advance, and likely course of action), (4) the depth of OPFOR advance during the attack, and (5) friendly force location with respect to the OPFOR and the MBA.

The values for each measure were collected on all pre- and post-test exercises and were presented to four expert military judges. The judges were retired officers, whose ranks ranged from lieutenant colonel to brigadier general. Each had extensive experience in combat or combat modeling. Based on the measures of mission objectives obtained for each exercise, the judges

gave an overall mission accomplishment score to each set of battle outcomes. Scores could range from 100, indicating perfect mission accomplishment, to 0, indicating total failure.

ARTEP Ratings. During pre- and post-test CATTS exercises, ratings of BCG performance were obtained from CATTS controllers, from the players themselves, and from player-controllers (company commanders and FIST). Items were based on a modified list of tasks taken from the command group/staff module of ARTEP 71-2, (Army Training and Evaluation Program . . . 71-2, 1977). These tasks (Appendix C) were selected on the basis of previous ARI research (Barber and Kaplan, 1979; Kaplan and Barber, 1979; Barber and Solick, 1980; and Thomas, Barber, and Kaplan, 1983). Ratings were obtained from a one to nine scale, where one corresponded to poor and nine to outstanding performance. In addition to rating group performance, the controllers also rated the performance of the principal group members.

RESULTS AND DISCUSSION

In general, battalion command group performance improved significantly from pretest to post-test exercises on all measures. The following provides a detailed description of the results for each performance measure.

Information Flow

The information-flow questionnaire measured three stages in the process of communication: (1) communication from brigade to battalion, (2) communication within the battalion command group and (3) communication from battalion to company. Prior to analysis of the information flow scores, the split-half reliability of the information-flow questionnaire based on a sample of 46 questionnaires was calculated and found to be a respectable .82. Further, the controllers' identification of required information was validated by the receivers' (players' and player-controllers') ratings of the importance of the items. The mean importance rating of all required items was 7.1, approximately midway between 5 (moderately important) and 9 (essential).

Table 3 presents the percentages of information received in each stage of communication. One-tailed t-tests were applied to the matched pretest and post-test scores for each of the five command groups. In all three stages the communication scores increased significantly from pretest to post-test. The greatest improvement (25%) occurred in the reception of information by the company commanders.

The controllers were also asked to observe the operations order briefings provided to the company commanders. Each controller was asked to record the number of required items briefed, to rate how good the briefings were, and to comment on the briefings. An average of 13.2 items were briefed in the pretest versus 17.6 items in the post-test, an increase that was not statistically significant. Further, the average rating of the briefings

Table 3

Effect of Training with Feedback
on the Mean Percent of Required Information Received

Stage of Communication	Pretest	Post-test	t
Brigade to Battalion	85	89	2.21*
Within the Bn Cmd Grp	51	66	4.97**
Battalion to Co Cdrs	42	67	4.96**

* $p < .05$ **$p < .01$ $df = 4$ one-tailed test for matched pairs

increased from 4.5 (less than good) in the pretest to 5.9 (half way between good and very good) on the post-test. This improvement was statistically significant at the .01 level ($t = 3.96$). Finally, the controllers commented on briefings that were particularly good or particularly bad. The most frequent comments (46%) referred to the completeness of the presentation. Typical unfavorable comments were that presentation was incomplete, sketchy, or omitted important information. Favorable comments were that it was thorough and covered most pertinent information. The next largest category (28%) concerned delivery. Critical comments were that the presentation was rushed, hurried, unsure, or unclear. Favorable comments were concise, clear, or good delivery. Some comments (8%) mentioned organization, saying the presentation was either poorly organized or well organized. In every category the percentage of favorable comments increased from the pretest to the post-test. Seventy-six percent of all the comments on the post-test were favorable, compared to 27% in the pretest.

Probes

Each probe was designed to elicit information-processing by at least one staff member, so such behavior could be sampled during mission execution. Some probes were more effective than others in eliciting these behaviors. Probes that were not responded to more than 40% of the time were consisidered ineffective and were excluded from further analysis. The remaining probes were scored correct or incorrect depending upon the completeness and accuracy of the responses, as assessed by the controllers who administered the probes. Percent of probes correct was calculated for each staff section (S1, S2, S3, S4, FSO) on each exercise day. These percentages were cast into a repeated measures ANOVA, to determine whether the command groups improved from pretest to post-test. It was found that the command groups' responses improved ($F_{1,16} = 4.68$, $p < .05$) as indicated in Table 4 and Appendix D. No significant difference among command groups (units) was observed in responses to probes. There was, however, a significant difference in percent correct between staff positions ($F_{4,16} = 4.03$, $p < .05$), where all staff positions performed better than the FSO (HSD = 4.23, $p < .05$) as indicated in Table 5 and Appendix D. This latter finding could mean either that FSO performance was inferior to that of other staff positions or that the FSO probes were more difficult than the others.

Mission Accomplishment

The degree to which judges agreed on their ratings of battlefield mission accomplishment was determined by correlating these scores among the four judges as indicated in Table 6. There was very high agreement among Judges 1, 2, and 3, and moderately high agreement between these judges and Judge 4. Since there was generally good agreement among judges' ratings, all mission accomplishment scores (Table 7) were included in an analysis to determine if these scores improved as a function of CATTS training.

Table 4

Mean Percent of Probes Correct
on Pretest and Post-test by Units

Unit	Pretest	Post-test
1	48	92
2	64	78
3	63	83
4	61	85
5	80	64
Mean	63	80

Table 5

Mean Percent of Probes Correct
on Pretest and Post-test by Staff Positions

Position	Pretest	Post-test
S1	87	90
S2	72	86
S3	77	87
S4	51	79
FSO	30	60
Mean	63	80

Table 6

Interrater Agreement on Mission Accomplishment Scores

	J_1	J_2	J_3	J_4
J_1	--	.951	.995	.757
J_2		--	.942	.748
J_3			--	.751
J_4				--

Table 7

Mean Mission Accomplishment Scores
for Units and Days

Unit	Pretest	Post-test
1	66.50	24.50
2	23.75	80.50
3	24.25	70.75
4	52.75	80.00
5	41.25	79.75
Mean	41.70	67.10

A 2 X 5 repeated measures ANOVA indicated that battalions tended to improve on battlefield mission accomplishment from pre- to post-test ($F_{1,2}$ = 14.79, $p < .05$). It was also determined that some battalions performed better on the simulated battlefield regardless of exercise day as indicated by a significant unit effect ($F_{4,12} = 3.61$, $p < .05$). Finally, as indicated in Table 7, one battalion (Unit 1) performed more poorly on the post-test than on the pretest. This is reflected in a significant unit by exercise day interaction* ($F_{4,12} = 16.47$, $p < .001$). The source table for this analysis appears in Appendix E.

ARTEP Ratings

Prior to testing for changes in ARTEP ratings from pretest to post-test, the ratings were analyzed to determine if the raters were (a) discriminating among items and (b) in agreement in their ratings of ARTEP performance.

Item Discrimination. Ratings of ARTEP performance made by controllers, players, and player-controllers were correlated between items for each rater and then combined for each rater group. These inter-item correlations were significant ($p < .01$) for all comparisons. The median interitem correlations for controllers, players, and player-controllers were .80, .71, and .65, respectively, indicating that raters typically did not discriminate among the various aspects of player performance on ARTEP tasks. That is to say, if command group performance was rated high on one ARTEP task, performance on all other ARTEP tasks tended to be rated high as well. Therefore, in subsequent data analyses an average ARTEP rating was used for each rater on each exercise day. This was achieved by calculating an arithmetic mean across all ARTEP items for each rater on each day.

Inter-rater Agreement. The degree to which observers agreed in their ratings of ARTEP performance could only be determined for controllers, since too little data was available from players and player-controllers. Inter-rater agreement was determined by correlating the average ARTEP ratings made by controllers on each exercise day. The highest agreement was between the chief controller and the S3 controller ($r = .90$). The S1 and S4 were also in high agreement in their ratings ($r = .88$). The lowest degree of relationship between ratings were between the FSO and the S1 controller ($r = .25$), and between the FSO and the S4 ($r = .29$). These results may reflect different perspectives on the exercises. For example, the chief controller and the S3 both observed from the brigade control station in the control room, the S1 and S4 sat next to each other at another station, and the FSO observed from a third station. Overall, the median correlation among raters was .63, indicating a moderately high level of agreement among controllers in their ratings of command group performance.

Command Group Performance Change. Separate analyses were performed on ARTEP ratings for each group of raters to determine if post-test performance was perceived as being higher than pretest performance. Since controllers

*The unit effect is also in part attributable to the low mission accomplishment score for Unit 1 on the post-test. The average score for this unit across days was lower than those for other battalions.

observed all exercises, more detailed analyses of their ratings was possible. All ratings of BCG performance on the set of ARTEP tasks were averaged across items for each rater on each exercise day. Controller ratings of BCG performance were analyzed in a 2 X 5 ANOVA to determine any pre- to post-test differences in ratings of command group performance. As indicated in Table 8 and verified by the analysis in Appendix F, command group performance was rated higher on the post-test exercises ($F_{1,52} = 93.90$ $p < .0001$), but no command groups were rated significantly higher than any others. Also there was no significant exercise day-by-unit interaction. That is to say, that controllers did not perceive any difference in the magnitude of the increase in ARTEP performance from pre-to post-test for any unit compared to the others.

Since differences in player and player-controller ratings between units and unit-by-day interactions were not of interest, only potential differences in ratings from pre- to post-test were analyzed using dependent t-tests. As shown in Figure 2, players perceived an increase in their ARTEP performance from the pretest to the post-test exercises ($t = 2.44$, $df = 4$, $p < .05$). Player-controllers also indicated that BCG ARTEP performance increased as a result of CATTS training as shown in Figure 2 ($t = 6.61$, $df = 4$, $p < .01$). In conclusion, all three groups of raters indicated that command group performance increased significantly after CATTS training with diagnostic feedback.

Ratings of Individual Staff Members. In addition to rating command group performance, controllers also rated the overall performance of each individual staff member whom they observed during CATTS exercises. All ratings were transformed to z-scores for each observer, to control for rater response bias. These data were then analyzed to determine if there were differences in performance between command groups, or differences in performance on the pre- versus the post-test. These ANOVA's are given in Appendix G. No significant differences in performance were observed between staff sections e.g., S3 players performed no better or worse than other group members. In addition, there was no significant difference in performance among members of different command groups. Finally, controller ratings of individual staff members did increase signficantly from the pretest exercise to the post-test ($F_{1,20} = 131.82$, $p < .001$). This increase in the ratings of individual performance is consistent with the pre- to post-test increase in command group ARTEP ratings described above.

Comparison With Previous Research

A previous investigation (Thomas, et al., 1983) obtained measures of performance similar to those described above, but did not employ the diagnostic and feedback procedures used in the present research. In that investigation, five battalion command groups participated in four-day CATTS exercises, and, as in the present experiment, ARTEP ratings and simulated battle performance measures were collected on the first and last exercise

Table 8

Mean Controller Ratings of Units
for Pre- and Post-test Exercises

Unit	Pretest	Post-test
1	3.80	5.54
2	3.80	6.77
3	4.16	6.21
4	4.29	7.23
5	3.94	6.20
Mean	4.00	6.39

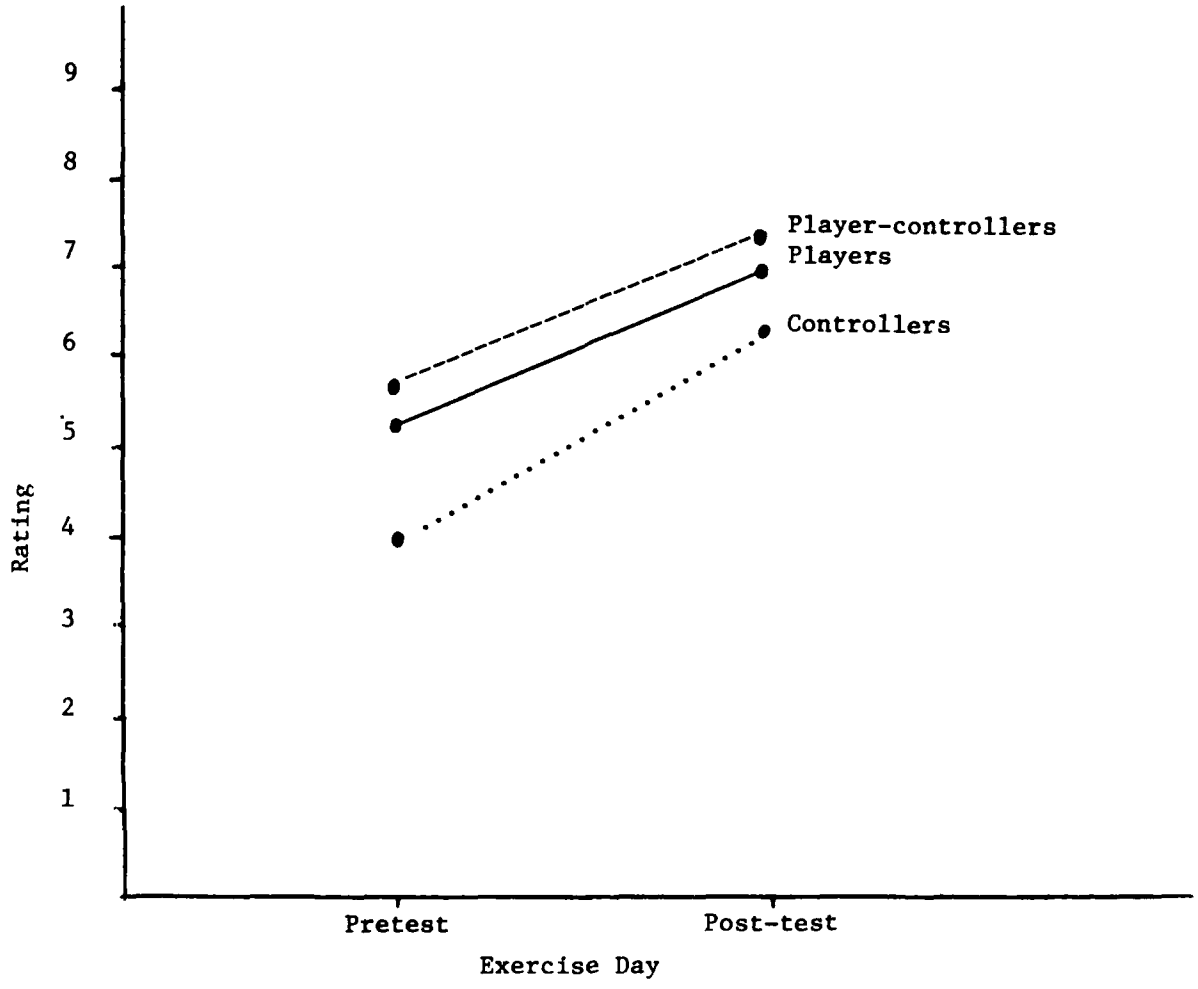


Figure 2. Mean ARTEP Performance Ratings by Controllers, Players, and Player-controllers on Pre-and Post-test Exercises.

days. Information-flow questionnaires were administered on the first and third days. Probes were not used, because the probe system had not been developed yet.

The statistical significance of the results obtained in the present and previous research is summarized in Table 9. As shown in the column headed "With Feedback," significant increases were observed for all performance measures after CATTS training supplemented by the diagnostic and feedback system. Without this feedback, however, CATTS training did not produce a significant increase in information flow scores. Nor did simulated battle performance improve without feedback. The players' and controllers' ratings of ARTEP performance did increase significantly, but the player-controllers' ratings did not.

In the previous experiment, feedback was limited to after-action reviews conducted by the players themselves. In the present experiment, the players received feedback about all the performance measures that were used. Although various differences between the two experiments may have contributed to the differences in results, it seems likely that the consistent increases in performance observed in the present research were due in large part to the diagnostic feedback system that supplemented the CATTS training.

CONCLUSIONS

Command group performance, as assessed by all measures developed in previous research and refined for the current experiment, improved significantly from the first day to the fourth day of CATTS training. All staff members improved in their ability to transmit and receive information during the planning phase as measured by the information flow questionnaire, and the execution phase as measured by probes. Battlefield mission accomplishment scores also improved as did ratings of command group performance by the players, controllers, and player-controllers, and ratings of individual member performance by the controllers.

In contrast, a previous experiment (Thomas, Barber, and Kaplan, 1983), in which feedback was limited to a brief after-action review at the end of each day, showed little evidence of improvement over a four-day exercise. There was no significant improvement in information flow scores or in battle simulation outcomes. The players' self-ratings and the controllers' ratings of command group performance did increase, but the player-controllers' ratings did not. Probes were not used.

In conclusion, BCG's improved on all measures of performance when diagnostic feedback was provided in addition to CATTS training. For these results to be generalized to ARTBASS, training exercises must be conducted in a manner similar to the procedures used in the current research. In particular, procedures for performance measurement should be implemented, and feedback sessions conducted by well-trained personnel should be included as part of the standard ARTBASS exercise.

Table 9

Improvement in Command Group Performance
With and Without Feedback
(Statistical Significance)

Performance Measure	With Feedback	Without Feedback
Information Flow		
Brigade to Battalion	$p < .05$	Not Significant
Within the Bn Cmd Grp	$p < .01$	Not Significant
Battalion to Co Cdrs	$p < .01$	Not Significant
Probes	$p < .05$	Not Used
Simulated Battle Performance	$p < .05$	Not Significant
ARTEP Ratings		
Players	$p < .05$	$p < .01$
Controllers	$p < .0001$	$p < .001$
Player-controllers	$p < .01$	Not Significant

REFERENCES

Army Training and Evaluation Program for Mechanized Infantry/Tank Task Forces, No. 71-2, Washington, D.C. Headquarters, Department of the Army, 17 June 1977.

Barber, H.F. and Kaplan, I.T. Battalion Command Group Performance in Simulated Combat. ARI Technical Paper 353, March 1979.

Barber, H.F. and Solick, R.E. MILES Training and Evaluation Test, USAREUR: Battalion Command Group Training. ARI Research Report 1290, June 1980.

Kaplan, I.T. and Barber, H.F. Training Battalion Command Groups in Simulated Combat: Identification and Measurement of Critical Performance. ARI Technical Paper 376, June 1979.

Thomas, G.S., Barber, H.F., and Kaplan, I.T. The Impact of CATTS System Characteristics on Selected Measures of Battalion Command Group Performance. ARI Technical Report 609, January 1984.

Thomas, G.S. and Cocklin, T.G. A Model of Mission Accomplishment in Simulated Battle. ARI Technical Report 599, November 1983.

APPENDIX A
First Page of an Information-Flow Questionnaire
Information Flow

Position _____ Date _____

Answer all questions. Circle the letter before the answer that you believe is correct. Do not guess. If you do not know the answer, answer "unknown". Some questions may cover information to which you did not have access.

Indicate how important it is for you to know each item by assigning it a number from 1 to 9, according to the following scale:

1	2	3	4	5	6	7	8	9
useless				moderately important				essential

1. The ADA weapons control status is _____.

   a. free
   b. tight — Importance _____
   c. hold
   d. unknown

2. Attack helicopters available to your TF for planning purposes is _____.

   a. one platoon
   b. two platoons — Importance _____
   c. one company
   d. unknown

3. Your TF has been instructed to go to BP _____ after passage of lines.

   a. 12
   b. 14 — Importance _____
   c. 8
   d. unknown

4. FM radio listening silence will be imposed until _____.

   a. lifted by TF headquarters
   b. reaching the SP — Importance _____
   c. reaching the RP
   d. unknown

5. The minefields and obstacles emplaced in your sector by the 201st ACR _____ by the engineers.

   a. are planned
   b. are in the process of being verified — Importance _____
   c. have been verified
   d. unknown

Controller _____

APPENDIX B
Example of a Probe
PROBE/OBJECTIVE MEASURE

Event	Insertion Time	Expected Response	Actual Response
Request SITREPS	Desired _____	S3 should examine records and provide Bde with situation report	
	Acutal _____		
		From Whom Bn S3	
		To Whom Bde S3	
		Timeliness Immediate	
		Completeness Front line trace friendly losses probability of completing mission others	
Comments:		Accuracy Compare with computer results 15 min prior	

APPENDIX C
Controller ARTEP Rating Form
Performance Estimate

Position _____ Date _____

Indicate how well you think the command group can perform tasks 1 to 5 below. Also rate the two overall performance items. Assign each task/item a number from 1 to 9, according to the following scale:

1	2	3	4	5	6	7	8	9
Poor		Fair		Good		Very Good		Outstanding

_____ 1. Gather and analyze required information. Includes: (a) analyze mission, (b) determine what information is available and what additional information is required, (c) determine what information sources are available, (d) gather all available information and request additional information as needed.

_____ 2. Develop a plan based on mission and modify it as required by events. Includes: (a) determine friendly capabilities and limitations, request additional assets if needed, (b) estimate enemy capabilities and likely courses of action, (c) identify key terrain, (d) select battle position/routes to objectives, (e) identify critical place, (f) develop and compare courses of action, (g) individual staff planning for communications, intelligence, operations, admin/log, fires, and (h) coordinate with other staff members.

_____ 3. Communicate/coordinate. Includes: (a) issue a warning order, (b) disseminate plans and orders, and (c) disseminate combat information and intelligence to higher and lower.

_____ 4. Implement plan. Includes: (a) concentrate/shift combat power, (b) reinforce terrain.

_____ 5. Supervise combat operations. Includes: (a) compare battlefield events with current order and concept of operations, (b) determine that a new course of action is necessary, and (c) determine that a change in implementation is necessary.

_____ Overall, how well did the command group perform its tasks?

Overall, how well did the staff members perform their own particular tasks?

_____ BC
_____ S1
_____ S2
_____ S3
_____ S4
_____ FSO
_____ ALO

APPENDIX D
ANOVA Tables for Probe Analyses

Table D-1

Source of Variance	Sum of Squares	DF	Mean Square	F	p
Day	3612.5	1	3612.5	4.68	<.05
Unit X Day	4739.6	4	1184.9	1.54	NS
Position X Day	1360.0	4	340.0	.44	NS
Residual	12342.4	16	771.4		

Table D-2

Source of Variance	Sum of Squares	DF	Mean Square	F	p
Position	11898.32	4	2974.58	4.03	<.05
Unit	73.52	4	18.38	.02	NS
Residual	11814.48	16	738.41		

APPENDIX E

Repeated Measures ANOVA on Mission Accomplishment Scores

Table E-1

Source	Sum Squares	Degrees of Freedom	Mean Squares	F	p
Judges	1034.2	3			
Day	6451.6	1	6451.6	14.79	< .05
Unit	2500.9	4	625.2	3.61	< .05
Unit X Day	12291.7	4	3072.9	16.47	< .001
Error Day	1309.0	3	436.3		
Error Unit	2075.6	12	172.96		
Error U X D	2238.8	12	186.56		

APPENDIX F

ANOVA Table for Controller Ratings of
Command Group Performance

Table F-1

Source of Variance	Sum of Squares	DF	Mean Squares	F	p
Raters	5.204	6			
Day	95.930	1	95.930	93.90	<.0001
Unit	8.569	4	2.097	2.10	NS
Unit X Day	3.942	4	.965	.97	NS
Residual	53.124	52	1.022		

APPENDIX G

ANOVA Tables for Controller Ratings of Individual Staff Member Performance

Table G-1

Source of Variance	Sum of Squares	DF	Mean Square	F	p
Unit	.656	4	.164	1.30	NS
Staff	.751	5	.150	1.19	NS
Residual	2.523	20	.126		

Table G-2

Source of Variance	Sum of Squares	DF	Mean Square	F	p
Day	3.791	1	3.791	131.82	.0001
Unit X Day	.157	4	.039	1.36	NS
Staff X Day	.145	5	.029	1.01	NS
Residual	.575	20	.029		